Quantidade de óbitos pela Covid-19 no Brasil entre 17/03/20 a 15/05/20: curva de tendência e projeções até 15/06/20

Alex Dias Gonsales {alexdiasgonsales@gmail.com}

23/05/2020

1 Introdução

Este trabalho faz uma análise do número de óbitos pela Covid-19 registrados no Brasil no período de 17/03/2020 a 13/04/2020 e efetua uma projeção para os 30 dias seguintes (até 15/05/2020). Após passados os 30 dias faz uma comparação entre a projeção efetuada e os dados reais ocorridos até essa data. Para finalizar é feita mais uma projeção para os 30 dias seguintes (até 15/06/2020). Quando chegar essa data será feita uma nova comparação e publicação desses novos resultados.

Dois objetivos principais permeiam este estudo:

1) Mostrar um exemplo prático de aplicação de curvas de tendência e equações de regressão;
2) Contribuir para uma compreensão básica do comportamento da curva de óbitos da Covid-19 no Brasil.

Em um próximo texto pretende-se escrever um tutorial passo a passo mostrando os detalhes da execução deste trabalho, facilitando para quem quiser replicar esta análise ou fazer outras de mesmo cunho.

2 Descrição do Método Utilizado

No dia 13/04/2020 foi acessado o site do Ministério da Saúde e baixado o arquivo que contém os dados relativos à quantidade de casos e número de óbitos pela Covid-19 registrados no Brasil. Esses dados foram importados em uma planilha eletrônica através da ferramenta Calc do pacote LibreOffice versão 6.4.3.2.

A partir desses dados foram criados alguns gráficos com respectivas curvas de tendência e equação de regressão. A partir dessas equações foram efetuadas as projeções futuras para 30 dias seguintes (até 15/05/2020). Após passados os 30 dias, acessou-se novamente o site do Ministério da Saúde baixando-se a planilha com os dados atualizados e comparando com as projeções que foram efetuadas.

Três conceitos importantes foram usados neste trabalho, os quais são explicados brevemente a seguir.

A **equação de regressão** é uma fórmula matemática que tem como objetivo mostrar qual é a possível equação matemática que gerou os dados em questão. A partir dessa equação pode-se fazer projeções para os valores futuros, caso a tendência de comportamento do evento estudado continue a mesma.

A **linha de tendência** é a exibição gráfica dos valores calculados utilizando-se a equação de regressão.

Já o **coeficiente de determinação**, também chamado de R-quadrado e denotado por R^2, é uma medida estatística que mostra quão próximo os dados reais estão dos dados calculados pela equação de regressão. Esse coeficiente varia de 0 a 1 e quanto mais próximo de 1 estiver significa que os dados estão mais próximos da curva de tendência especificada.

Neste trabalho utilizou-se para as curvas de tendência uma regressão do tipo polinomial de grau 3 pois esta mostrou um R^2 mais próximo de 1 quando comparado aos outros tipos de regressão disponíveis na ferramenta utilizada (linear, logarítmica, exponencial, geométrica e polinomial de grau 2).

As atividades executadas consistiram basicamente de três etapas. Na primeira etapa são criados os gráficos a partir dos dados disponíveis e obtidas as equações de regressão. Na segunda etapa são efetuadas projeções para 30 dias seguintes, utilizando-se as equações de regressão obtidas na etapa 1. Na terceira etapa, após decorridos os 30 dias, as projeções efetuadas são comparadas com os dados reais ocorridos.

2.1 Geração dos Gráficos e Equações de Regressão

A partir dos dados disponíveis até o dia 13/04 foram gerados alguns gráficos e obtidas as suas respectivas equações de regressão. Foram feitas quatro análises, cada uma com um período diferente de dados, conforme mostrado na Tabela 1. Cada análise gerou uma equação de regressão que será usada na etapa 2 para fazer as projeções para os 30 dias seguintes. O detalhamento do processo de geração dos gráficos e de como se obtém as equações de regressão não será mostrado aqui. O objetivo é somente mostrar as equações obtidas e as projeções que foram feitas a partir dessas equações, comparando os resultados. O detalhamento de todo processo será descrito em breve em outra publicação.

Em todas as análises utilizou-se como dia 1 (primeiro dia) a data do primeiro óbito registrado no Brasil, ou seja, 17/03/2020. Portanto, o dia 18/03 é considerado o dia 2, o dia 19/03 é o dia 3 e assim por diante. Pode-se notar que do dia 1 (17/03) até o dia desta análise em questão (13/04) existem 28 dias. Dentro desse período de 28 dias foram feitas 4 análises diferentes, cada uma com uma quantidade crescente de dias (Tabela 1) de forma a poder compará-las e verificar se a quantidade de dados (dias) teria influência sobre a qualidade das projeções efetuadas.

Tabela 1: Faixas de períodos analisados

Análise	Período Analisado	Quantidade de dias
Análise 1	17/03 a 23/03	7
Análise 2	17/03 a 30/03	14
Análise 3	17/03 a 06/04	21
Análise 4	17/03 a 13/04	28

Os gráficos gerados em cada uma dessas análises são mostrados nas figuras 1, 2, 3, 4 e 5. Em todos esses gráficos a equação de regressão considera a variável X (data) como o número de dias desde o primeiro dia considerado (17/03), conforme já foi explicado.

As figuras 1 e 2 mostram os gráficos da análise 1, sendo que a figura 1 mostra o eixo X em formato de data (para melhor visualização) e a figura 2 mostra o eixo X em formato de dia do evento. Este último formato é que deverá ser utilizado para aplicação das equações de regressão obtidas. No entanto, todos os demais gráficos neste trabalho foram mostrados somente em formato de data, para melhor visualização e para evitar colocar uma quantidade excessiva de imagens neste documento. Logo, quando for utilizar a equação de regressão para calcular algum valor, a data (variável x) deverá ser convertida para o número do dia do evento.

As equações de regressão obtidas em cada um dos gráficos foram as seguintes:

$$f(x)=0{,}0000x^3+0{,}7023x^2-0{,}1547x+0{,}7142 \quad \text{(eq. 1)}$$

$$f(x)=0{,}0161x^3+0{,}6013x^2-0{,}2430x+1{,}2137 \quad \text{(eq. 2)}$$

$$f(x)=0{,}0781x^3+0{,}6608x^2+6{,}5823x-7{,}1679 \quad \text{(eq. 3)}$$

$$f(x)=0{,}0580x^3+0{,}2473x^2-3{,}5356x+16{,}2981 \quad \text{(eq. 4)}$$

Para demonstrar a aplicação de uma dessas equações, vamos usar como exemplo o gráfico 4 e a sua respectiva equação de regressão (equação 4). Vamos calcular o total de óbitos acumulados para o dia 05/04, que é o vigésimo dia conforme explicado anteriormente. Portanto, na equação 4 vamos substituir o valor de x por 20 e resolver a equação:

$$f(x)=0{,}0580\cdot x^3+0{,}2473\cdot x^2-3{,}5356\cdot x+16{,}2981$$

$$f(x)=0{,}0580\cdot 20^3+0{,}2473\cdot 20^2-3{,}5356\cdot 20+16{,}2981$$

$$f(x)=0{,}0580\cdot 8000+0{,}2473\cdot 400-3{,}5356\cdot 20+16{,}2981$$

$$f(x)=464+98{,}92-70{,}712+16{,}2981$$

$$f(x)\approx 508{,}5$$

Isso resulta no valor 508, que é o valor aproximado do total de óbitos acumulados no dia 05/04 e que pode ser confirmado no gráfico da análise 4 (figura 5).

Uma observação pontual se faz necessária com relação à equação 1. Essa equação apresentou uma situação peculiar com relação ao coeficiente de x^3 pois este está com valor zero (até a quarta casa decimal). Como efeito prático essa equação acaba sendo equivalente a uma equação de grau 2, mas isto não faz diferença para o contexto deste trabalho. Este foi o único gráfico em que isto ocorreu, provavelmente por ser um conjunto com pouca quantidade de dados (apenas 7 dias). Em todos os outros gráficos a equação de regressão gerada foi de grau 3.

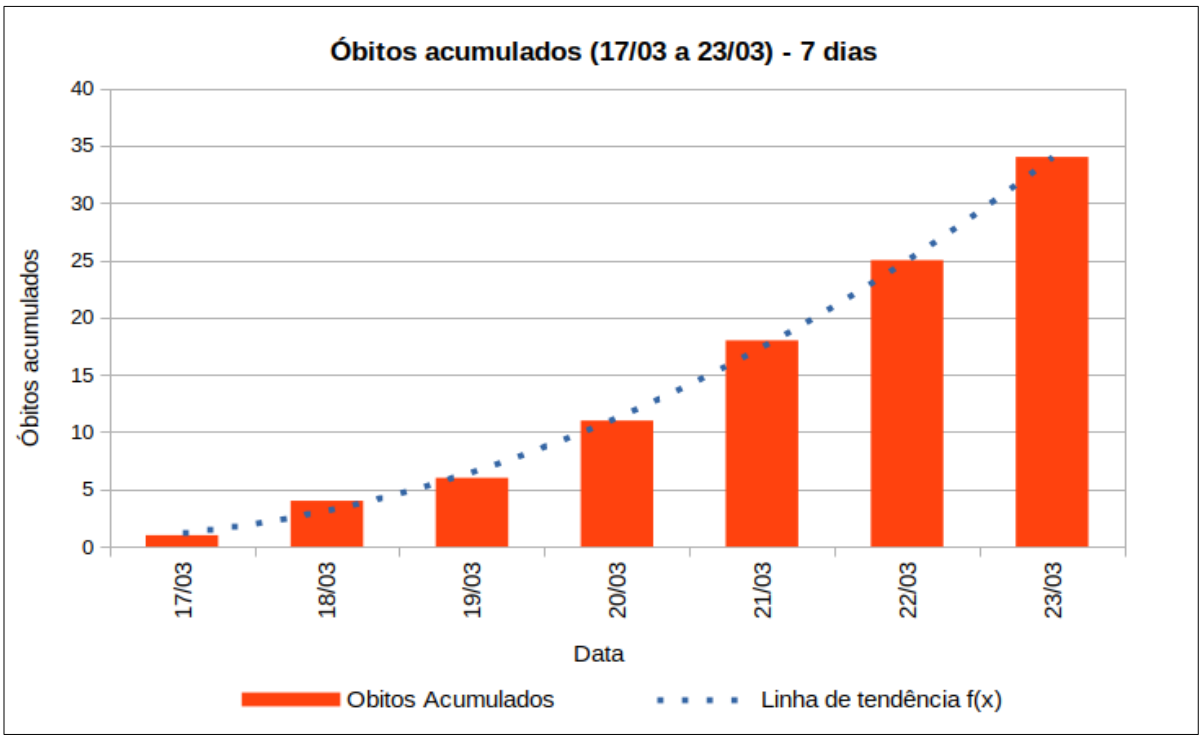

Figura 1: Análise 1 com eixo x em formato de data.

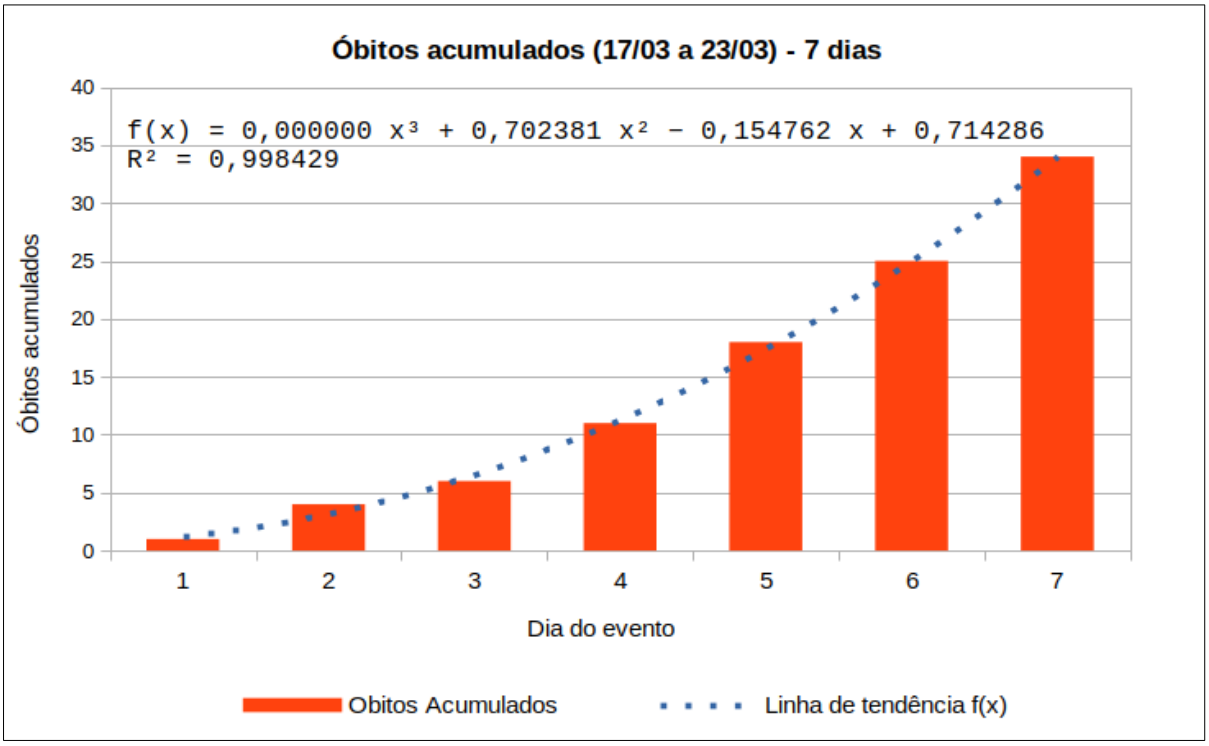

Figura 2: Análise 1 com eixo x em formato de dia do evento.

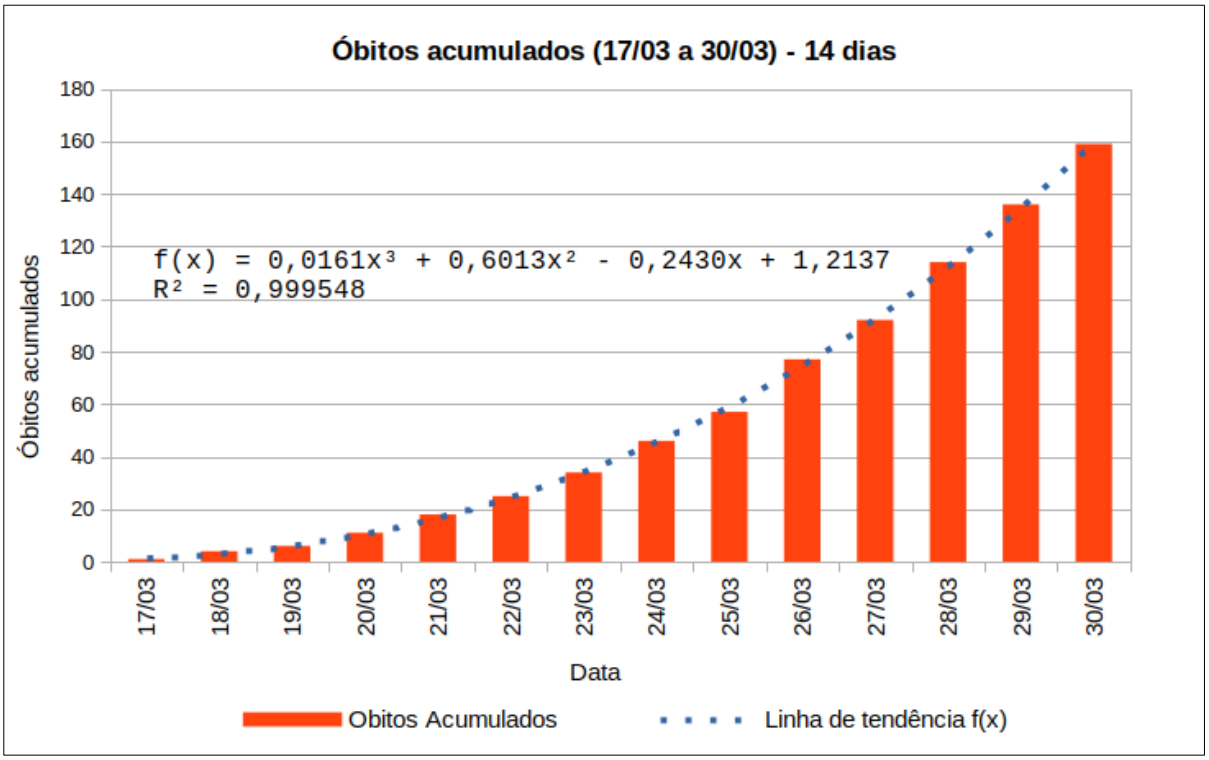

Figura 3: Análise 2.

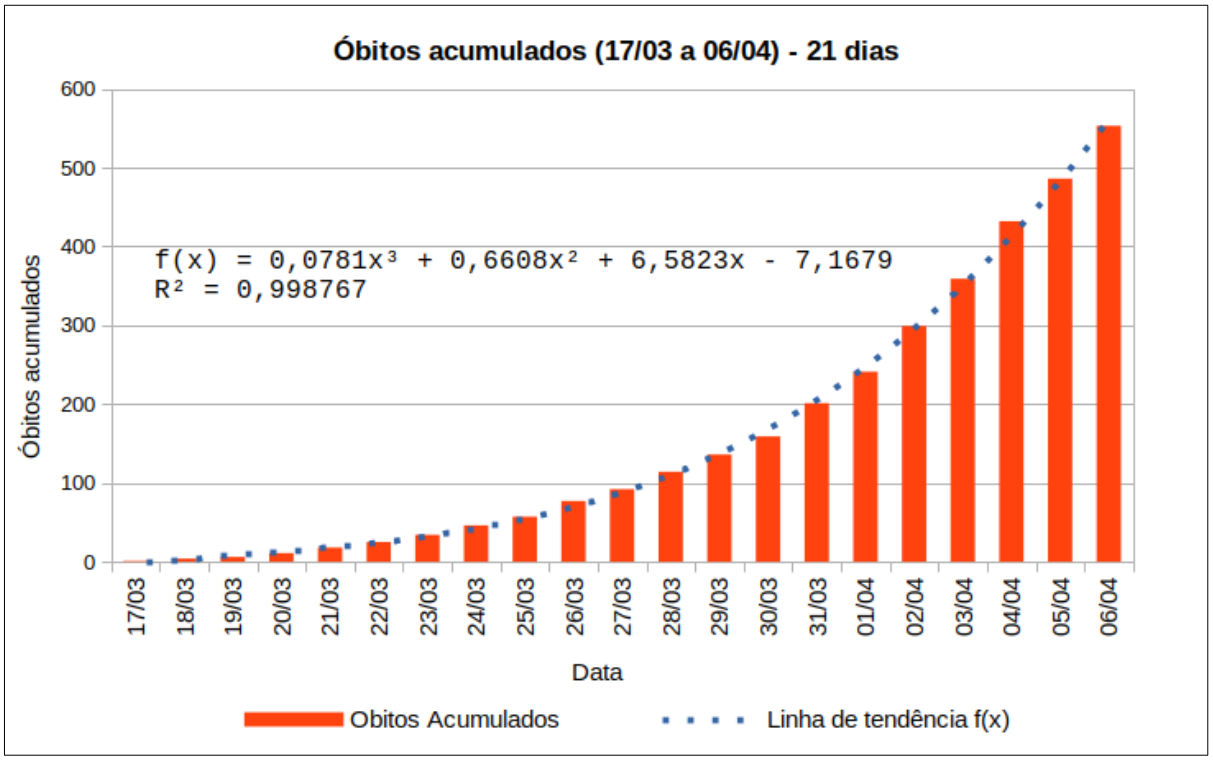

Figura 4: Análise 3.

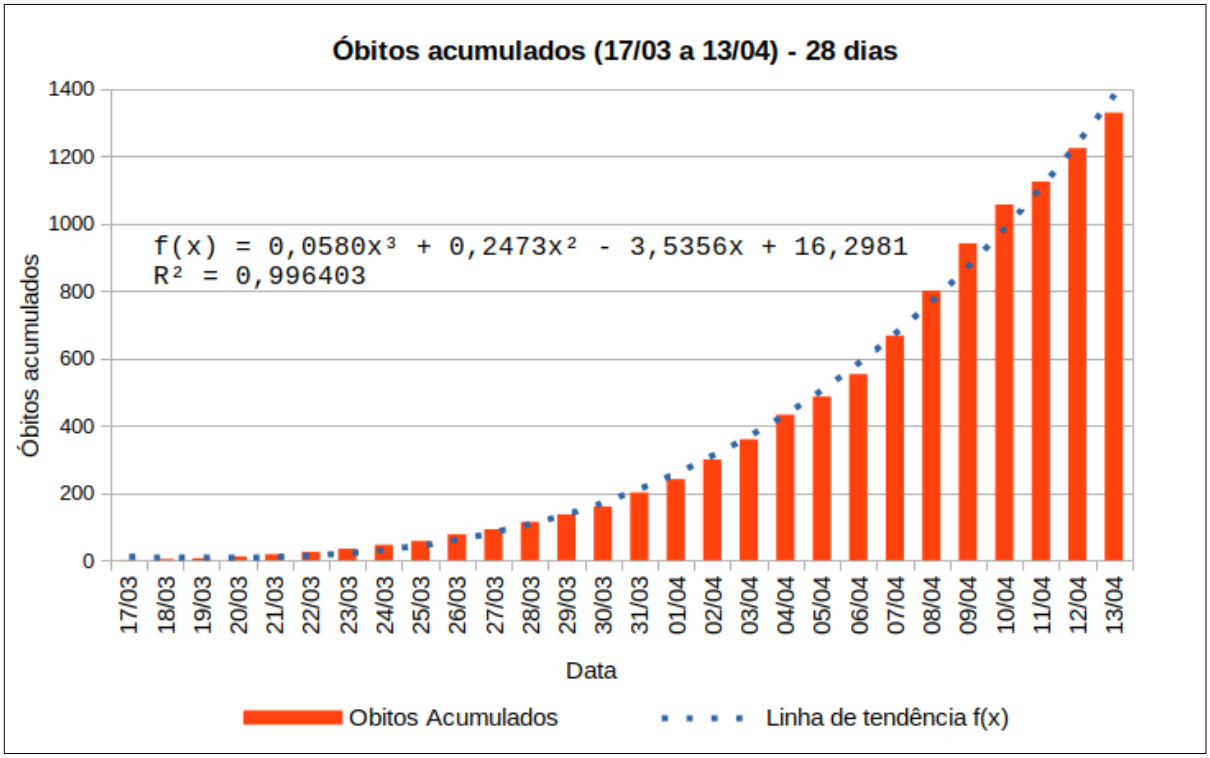

Figura 5: Análise 4.

2.2 Análise dos gráficos e projeções

Observando-se os gráficos gerados em cada uma das 4 análises é possível notar que a linha de tendência (pontilhado em azul) está muito próxima dos valores reais de óbitos registrados (barras vermelhas). Isso indica que as equações de regressão conseguiram aproximar-se dos valores reais ocorridos, quando se olha apenas o intervalo de cada análise em questão. Da mesma forma, os valores de R^2 estão muito próximos de 1 (tabela 2) o que também indica a proximidade dos valores reais com os valores obtidos pelas equações de regressão.

Tabela 2: Valores de R-quadrado

Análise	**Valor de R^2**
Análise 1	0,998429
Análise 2	0,999548
Análise 3	0,998767
Análise 4	0,996403

É sabido, no entanto, que quanto maior a quantidade de dados disponíveis sobre um sistema, mais fiel serão as análises que se pode fazer sobre tal sistema. Conforme mostrado na tabela 1, das quatro análises efetuadas, a análise 4 foi a que utilizou a maior quantidade de dados (28 dias) disponíveis até aquele momento (13/04/2020). Dessa forma, utilizou-se a equação 4 para se fazer uma projeção dos óbitos para os próximos 30 dias (até 15/05/2020). Essa projeção foi chamada de projeção 4, conforme pode ser vista na figura 6. Cabe salientar novamente que essa projeção só seria válida se as dinâmicas de letalidade do vírus e respectivo registro de óbitos continuassem as mesmas dos 28 dias considerados na análise 4.

Não é intuito deste trabalho tentar fazer qualquer previsão, mas apenas mostrar uma aplicação prática de alguns recursos da ferramenta Calc, tais como, gráficos, curvas de tendência, equação de regressão e R^2. Ao mesmo tempo, estamos fazendo um teste para verificar se a dinâmica da letalidade da Covid-19 no Brasil se mantém a mesma ao longo das semanas e meses.

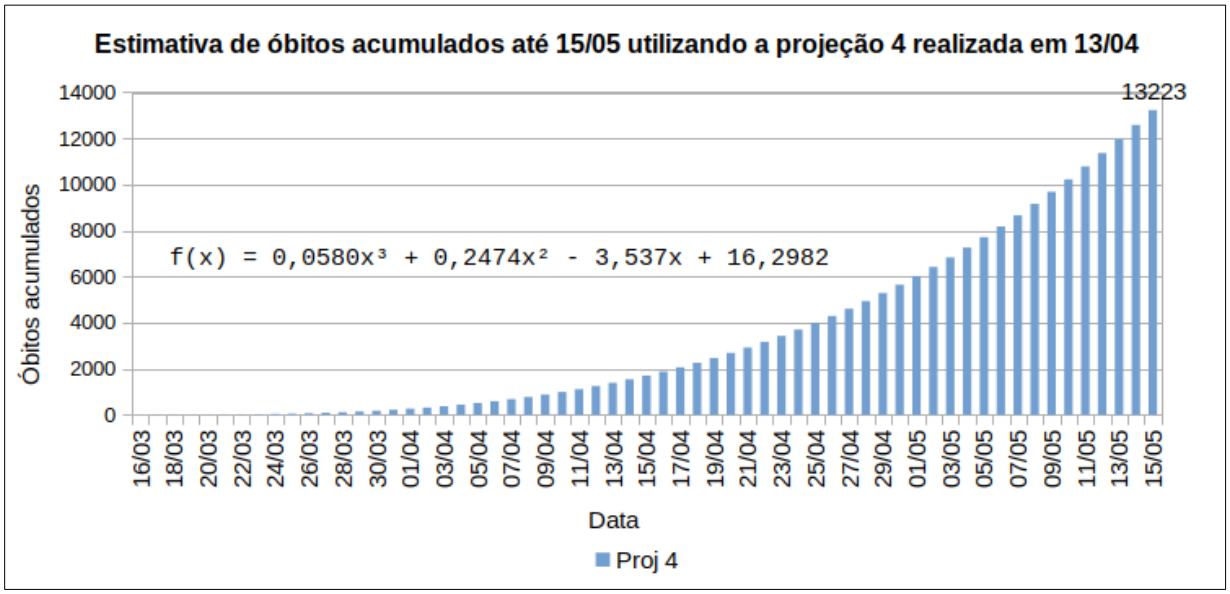

Figura 6: Projeção 4.

2.3 Comparando os resultados

Após passados 30 dias da realização dessa análise, baixou-se novamente o arquivo atualizado (até 15/05/2020) do Ministério da Saúde com os registros de óbitos, gerando-se o gráfico, conforme mostra a figura 7. A partir desses dados atualizados foi realizada uma comparação entre esses registros e a projeção 4 que havia sido realizada no dia 13/04 (fig. 6). O resultado pode ser visto na figura 8.

Observa-se que até o dia 05/05 as duas curvas estão muito próximas e a partir dessa data começa haver um distanciamento, para mais, dos registros de óbitos em relação à projeção efetuada, atingindo uma diferença de 12% (14.817 / 13.223) no dia 15/05. Não é intuito deste texto discutir os motivos dessa diferença, mas pode-se citar algumas possíveis causas bem óbvias:

1) A dinâmica de letalidade do vírus modificou-se;
2) O método de registro de óbitos foi modificado;
3) O tipo de regressão utilizado (polinomial de grau 3) neste modelo não foi suficiente para captar todas as informações do conjunto de dados disponíveis.

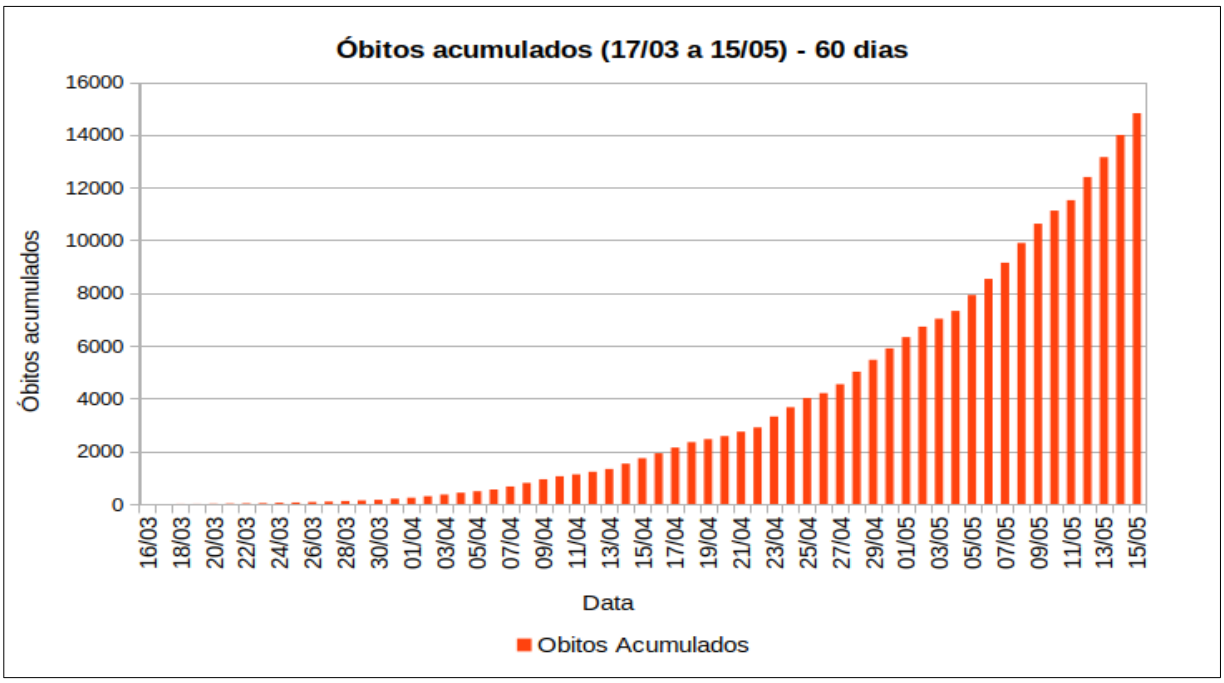

Figura 7: Registro de óbitos de 17/03 a 15/05 (60 dias).

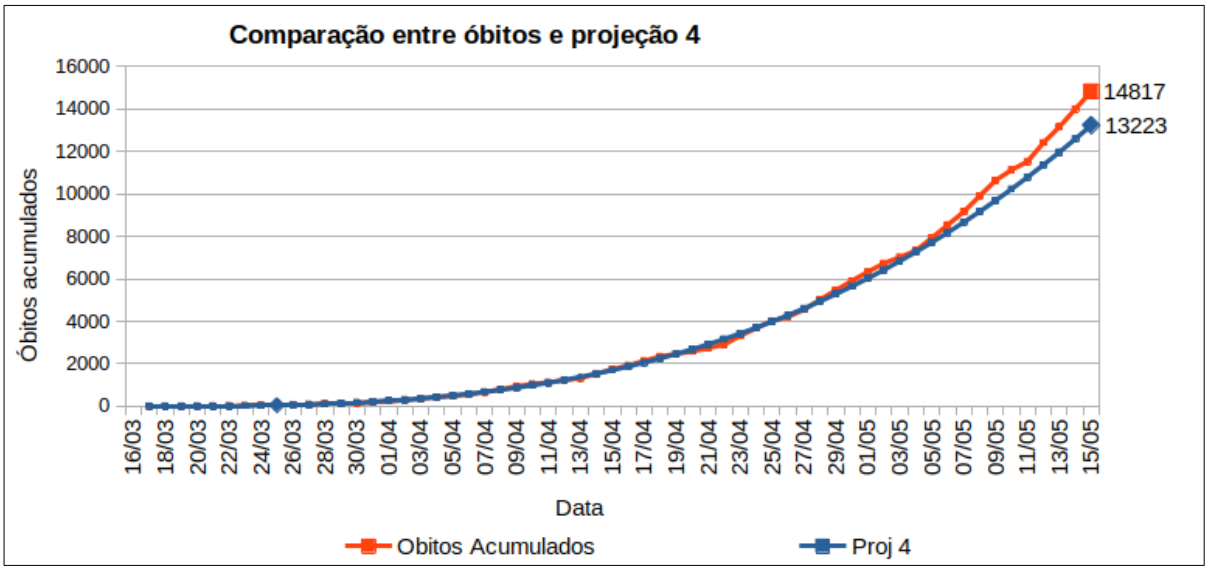

Figura 8: Comparação entre óbitos registrados e projeção 4.

Com intuito de ampliar este estudo, foram realizadas mais três comparações dos dados atualizados (até 15/05) com projeções utilizando-se as equações de regressão 1, 2 e 3 que foram geradas anteriormente. Como já foi dito antes, essas equações não deveriam dar melhores resultados, visto que a análise que possuía a maior quantidade de dados (28 dias) foi a análise 4 (equação 4). Os resultados das projeções 1, 2 e 3 são mostrados a seguir nas figuras 9, 10 e 11.

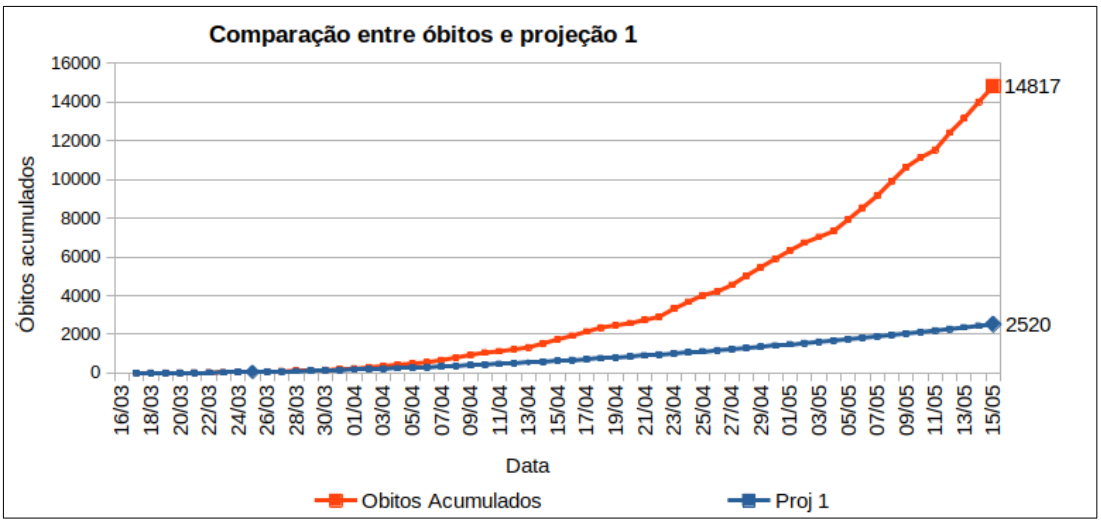

Figura 9: Comparação entre óbitos e projeção 1.

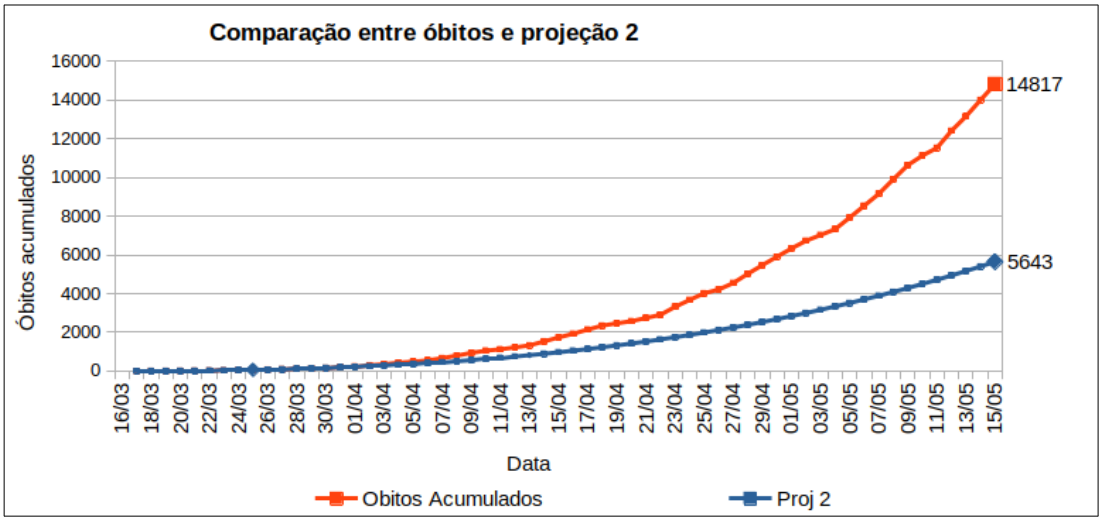

Figura 10: Comparação entre óbitos e projeção 2.

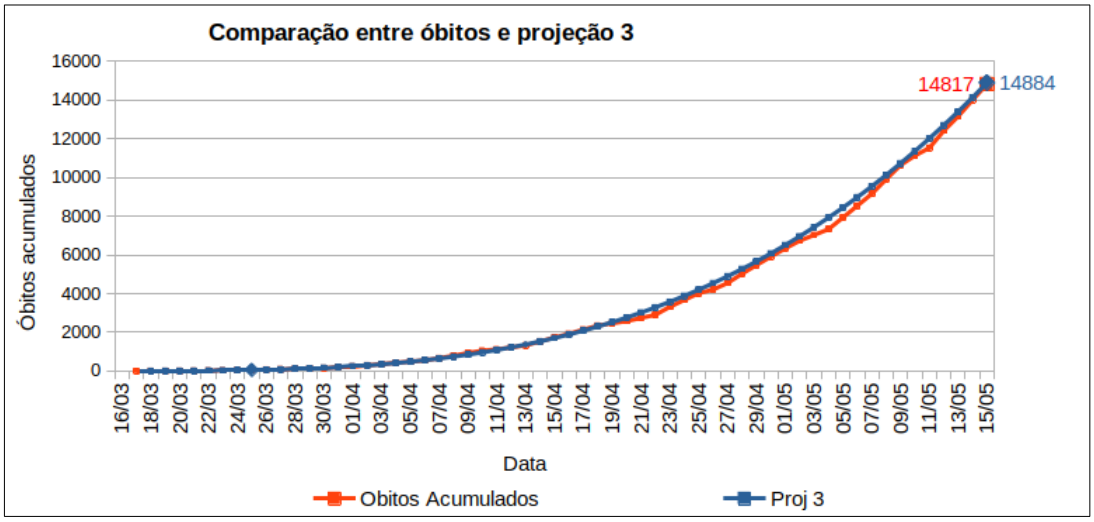

Figura 11: Comparação entre óbitos e projeção 3.

Como era de se esperar, as projeções 1 e 2, que foram construídas com as equações 1 e 2 respectivamente, as quais foram obtidas de conjuntos com poucas quantidades de dados (7 e 14 dias respectivamente), estiveram bem fora da realidade observada. Essa diferença é extrapolada no último dia da projeção (15/05). Observa-se que a projeção 1 (figura 9) apresentou uma diferença de 587% (14.817 / 2.520) e a projeção 2 (figura 10) apresentou um diferença de 262% (14.817 / 5.643).

Já as projeções 3 e 4, que utilizaram as equações 3 e 4 respectivamente, estiveram muito próximas da realidade. Conforme já mostrado anteriormente, a projeção 4 (fig. 8), que foi construída com a equação 4, a qual foi obtida a partir de um conjunto de dados com 28 dias, apresentou uma diferença de apenas 12% no último dia da projeção (15/05).

No entanto, algo curioso aconteceu com a projeção 3 (figura 11), que foi construída com a equação 3, a qual foi obtida a partir de um conjunto de dados com 21 dias, e que apresentou uma diferença de apenas 0,4% (14.817 / 14.884) no último dia da projeção (15/05). Ou seja, apesar da projeção 3 ter utilizado uma equação baseada em um conjunto de dados com 21 dias, portanto 7 dias a menos que a projeção 4, que tinha 28 dias, o final da sua curva esteve mais próxima dos dados reais do que a curva da projeção 4. No entanto, no dia 05/04, a projeção 3 apresentou uma diferença de 8% (7321 / 7920) em relação aos dados reais.

Acredita-se que a dinâmica da letalidade da Covid-19 e do método de registro dos óbitos sofre pequenas modificações ao longo do tempo, pois diversas variáveis estão envolvidas nesse processo, portanto, nenhuma equação matemática conseguirá modelar com precisão todos esses aspectos.

3 Novas Projeções

A partir dos novos dados atualizados (figura 7) e utilizando-se dos mesmos métodos anteriores, fez-se mais uma análise, com geração do gráfico, curva de tendência e equação de regressão, conforme mostra a figura 12 e equação 5 abaixo.

$$f(x) = 0{,}0930^3 - 2{,}1049\,x^2 + 37{,}9848\,x - 138{,}0167 \quad \text{(eq. 5)}$$

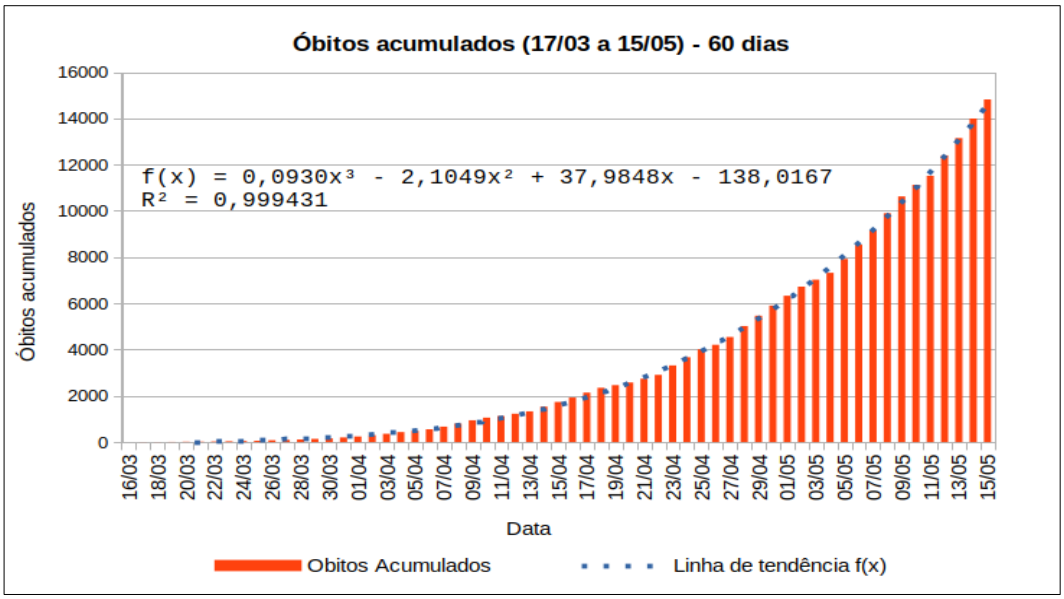

Figura 12: Registro de óbitos de 17/03 a 15/05 (60 dias).

A partir dessa equação foi realizada a projeção 5 e estendeu-se a projeção 4 até o dia 15/06/20, gerando-se o gráfico com essas duas projeções (figura 13). Observa-se que, para o dia 15/06, a projeção 4 estima um total de 45.452 óbitos e a projeção 5 estima um total de 56.014 óbitos.

Este trabalho foi finalizado no dia 23/05/2020 e está utilizando as informações de óbitos registrados até o dia 15/05/2020. A figura 14 mostra uma comparação dos dados reais até a data de 15/05 com as projeções 4 e 5. Quando chegar o dia 15/06/2020 os registros de óbitos serão atualizados e os dados da curva de óbitos será completada para devida comparação com as projeções 4 e 5.

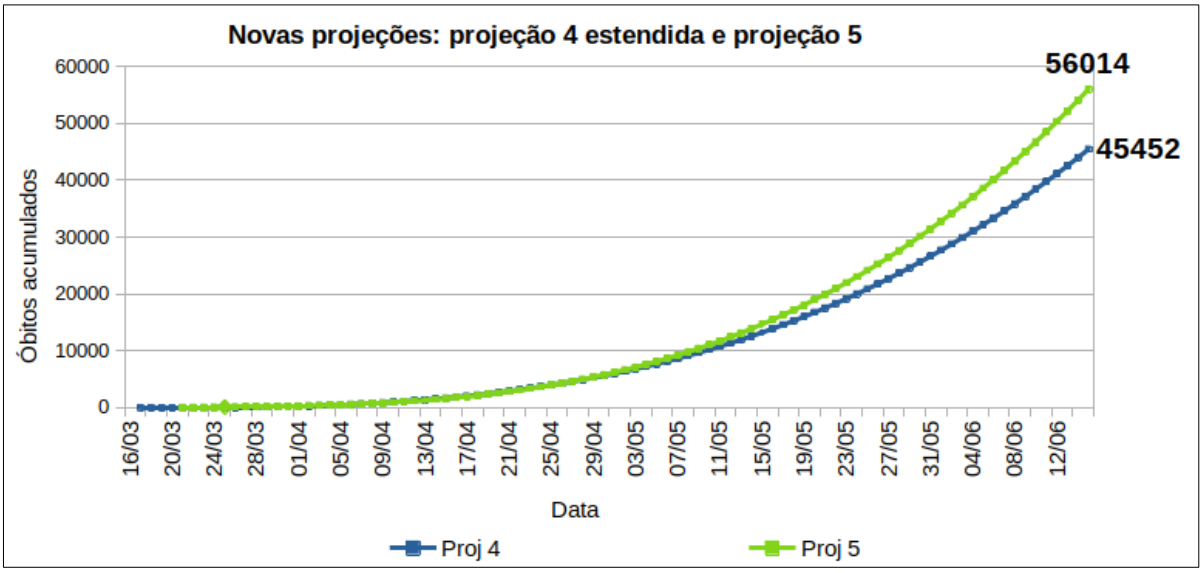

Figura 13: Novas projeções (projeção 4 estendida e projeção 5)

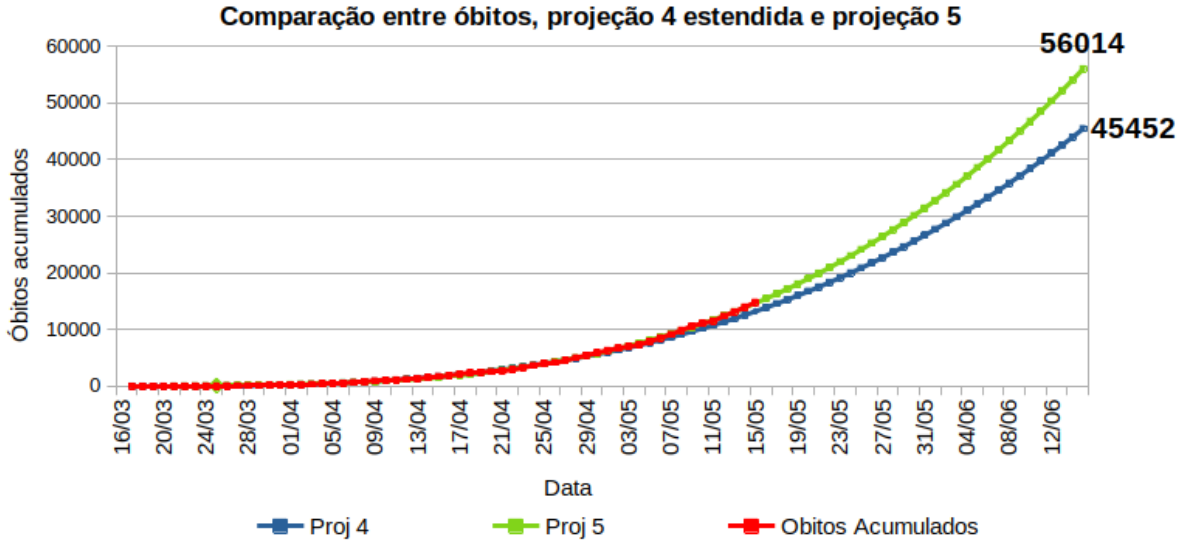

Figura 14: Novas projeções e dados reais até 15/05/2020.

4 Considerações Finais

Este trabalho mostrou um exemplo de utilização de curvas de tendência e equações de regressão, sendo que sua construção prática foi feita com uma planilha eletrônica, utilizando-se a ferramenta Calc do pacote LibreOffice. Estas mesmas técnicas podem ser aplicadas na ferramenta Excel da Microsoft. Acredita-se que este material possa servir de recurso auxiliar para professores e alunos de cursos relacionados à matemática, estatística e informática básica (planilhas eletrônicas).

A partir das análises feitas pode-se concluir que nos 30 primeiros dias analisados (a partir do primeiro registro de óbito pela Covid-19 no Brasil), que compreendem o período de 17/03/2020 a 13/04/2020, e os 60 primeiros dias que compreendem o período de 17/03/2020 a 15/05/2020, não se observou uma tendência de diminuição no número de óbitos.

Isso significa que, se estas projeções estiverem corretas e se a letalidade da Covid-19 se mantiver com a mesma dinâmica, daqui a 30 dias (em 15/06/2020) teremos um aumento expressivo no número de óbitos, passando de 14.817 para algo em torno de 50.000 (um aumento de cerca de 340%).

Com esses resultados espera-se contribuir para uma compreensão básica a respeito das curvas de crescimento no número de óbitos pela Covid-19 no Brasil e quem sabe ajudar na conscientização sobre os graves problemas que ainda teremos que enfrentar pelos próximos períodos.

5 Referências

Calc – LibreOffice. Disponível em http://pt-br.libreoffice.org.

Coronavírus Brasil. Disponível em http://covid.saude.gov.br